Time to act: a model for exploring and exploiting time-dependent instrumental contingencies

A time to plant,  
    And a time to pluck *what is* planted;

…

A time to gain,  
    And a time to lose;  
A time to keep,  
    And a time to throw away;

Ecclesiastes 4

Real-life context -- AD

* Harvesting a time-dependent resource, e.g. grapes for wine
* Debt
* When to ask your boss for a pay raise
* Temporal aspect of the decision to commit suicide

Empirical studies and formal approaches

In ancient activities of hunting and harvesting and in social interactions such as negotiations (Sinaceur and Neale 2005), courtship, and humor timing decides the success of our actions. Good timing is often not obvious and needs to be learned from experience. To learn these tasks, one typically develops a representation of a partially observed time-dependent process: movement of prey, fruit ripening or another person’s internal state. To the extent that such a process controls the reward associated with an action, we term it a *time-dependent contingency.*  [overview of empirical and formal work on temporal contingency learning]

Approaches to time-dependent contingencies in Pavlovian conditioning

[Inapplicability of TD: explain why the state space partition

[Seems like we need a brief review of TD(lambda) here to clarify that continuous time modeling has accomplished in the Pavlovian context]

In classical conditioning, an organism learns to predict the likelihood of a biologically salient outcome (i.e., the unconditioned stimulus) on the basis of one or more neutral stimuli that precede the outcome (i.e., conditioned stimuli). Reinforcement learning models of classical conditioning have long incorporated continuous-time representations of subjective value in order to characterize the predictive value of conditioned stimuli. One of the best developed is the temporal difference (TD) model which is based on the premise that organisms use information about the time between a CS presentation and receipt of the US to ….

[I self-plagiarized the following TD overview from our meta-analysis]

“Temporal difference (TD) models of animal learning, like RW, learn from prediction errors (Sutton & Barto, 1998), and describe an approach modeling prediction and optimal control. TD aims to predict all future rewards, discounting them over time:

Eq. 3: R(t)=r(t+1)+ γr(t+2)+ γ2r(t+3)+…+ γkr(t+k+1)

where r is future reward and γ is the temporal discount factor reflecting a preference for immediate over delayed rewards. Instead of waiting until all the outcomes are experienced, TD estimates future rewards by repeating the following algorithm in each learning episode (time step).

Eq. 4: V(t)🡨V(t) + α[r(t+1)+γV(t+1)-V(t)]

where α[r(t+1) + γV(t+1)-V(t)] is the prediction or *temporal difference error*, and r(t+1) + γV(t+1) takes the place of the remaining terms γr(t+2)+ γ2r(t+3)+…+ γkr(t+k+1).

To deal with the temporal distribution of predictive cues or response options, TD methods introduce the idea of *eligibility traces*. That is, only closely preceding (eligible) cues or actions are credited for reward or blamed for punishment.

TD provides a real-time account of learning, that RW or other trial-level models do not. A key area of divergence between RW and TD is that TD treats rewards themselves and the cues that predict them as, in principle, equivalent, insofar as they are both stimuli, which can invoke changes in the valuation of future rewards. Both conditioned cues and outcomes can influence value prediction, and can elicit prediction errors. This innovation provides an effective account of the learning of sequences of stimuli, as conditioned cues can come to operate as reinforcers in their own right (Dayan & Walton, 2012).”

The basic implementation of TD(

Although computational models of classical conditioning have yielded important insights into the representation of time in reward learning, there have been fewer attempts to understand how individuals learn an instrumental contingency in which the reward received depends on the timing of the response.

[Frank here]

[On instrumental tasks: no normative

* Clock – MH
* Frank’s model -- MH
  + +Uncertainty-driven exploration
  + –Limited representation
    - Value: can only get monotonic contingencies
    - Uncertainty: unable to pick a region of maximum uncertainty
  + -RTbest
  + – additive model combining learning & choice rule
* Opportunity cost
  + WTW – AD
  + Niv & R-learning, vigor, opportunity cost; Redish -- AD

Temporal basis

Fiorillo and colleages have suggested that the temporal representation of timing in midbrain dopamine neurons scales with the mean CS-US interval (learned over hundreds of trials) and that the precision of timing declines as intervals increase (Fiorillo, Tobler, & Schultz, 2003).

.

* Temporal representation of value, time as a high-dimensional action space – Ludvig, Kelsoe & Sutton (good but does not extend to instrumental paradigms – Moustafa 2008, propagation) -- MH

Model [for now paste the equations]

* Temporal basis representation of ACTION value, not the same as state value in TD
* Temporal basis for uncertainty – predicts where precisely one will explore
* Opportunity cost based on reward rate
* Value-uncertainty competition

Value representation:



Uncertainty representation:



Choice rule:







Context

Opportunity cost

*Time is the most valuable thing a man can spend.*

*Theophrastus*

Learning a temporal contingency does not yet guarantee optimal timing. That optimal timing has to also reflect the intrinsic value of time. This value can be thought of as the *opportunity cost*, or the alternative rewards that are foregone while waiting or performing an action. A good estimate of the opportunity cost is given by the long-term reward rate (Niv, Daw et al. 2005). Indeed, animals respond faster at higher reward rates (Herrnstein 1970).

Animals have evolved philogenetically conserved neural architecture for learning how to best time their actions. One critical part of this architecture are the dopaminergic circuits linking the midbrain with the basal ganglia.

Temporal basis

*The only reason for time is so that everything doesn't happen at once.*

*Albert Einstein*

Herrnstein, R. J. (1970). "On the law of effect." Journal of the Experimental Analysis of Behavior **13**(2): 243-266.

Niv, Y., N. D. Daw and P. Dayan (2005). How fast to work: Response vigor, motivation and tonic dopamine. NIPS.

Sinaceur, M. and M. Neale (2005). "Not All Threats are Created Equal: How Implicitness and Timing Affect the Effectiveness of Threats in Negotiations." Group Decision and Negotiation **14**(1): 63-85.